

Source and Accuracy Statement for the March and April 2000 CPS Microdata File for Child Support

SOURCE OF DATA

The data for this microdata file came from the March and April 2000 Current Population Survey (CPS). The Census Bureau conducts the CPS every month, although this file has only March and April data. The March and April surveys use two sets of questions, the basic CPS and the supplement.

Basic CPS. The monthly CPS collects primarily labor force data about the civilian noninstitutional population. Interviewers ask questions concerning labor force participation about each member 15 years old and over in every sample household.

CPS March and April 2000 Supplements. In addition to the basic CPS questions, interviewers asked supplementary questions in March 2000 about the economic situation of persons and families for the previous year. Of the housing units interviewed in March, about 45,774 were also interviewed in April. All household members 15 years of age and older that are a biological parent of children in the household from an absent parent were asked detailed questions about child support and alimony.

In March 2000, there were 4,393 household members eligible of which 1,386 required imputation of child support data. When matching the March 2000 and April 2000 data sets, there were 107 eligible people on the March file that did not match to people on the April file. Child support data for these 107 people were imputed. The remaining 1,279 imputed cases were due to nonresponse to the child support questions. Table 1 gives the sample sizes and the imputation rates by marital status.

Table 1. Sample Sizes and Imputation Rates

Marital Status	Sample Size	Imputed Cases	Rate
Married	1,016	285	28%
Widowed	70	23	33%
Divorced	1,597	500	31%
Separated	475	150	32%
Never Married	1,235	428	35%
Total	4,393	1,386	32%

Sample Design. The present monthly CPS sample was selected from the 1990 Decennial Census files with coverage in all 50 states and the District of Columbia. The sample is continually updated to account for new residential construction. To obtain the sample, the United States was divided into 2,007 geographic areas. In most states, a geographic area consisted of a county or several contiguous counties. In some areas of New England and Hawaii, minor civil divisions are used instead of counties. These 2,007 geographic areas were then grouped into 754 strata, and one geographic area was selected from each stratum.

About 50,000 occupied households are eligible for interview every month out of these 754 strata. Interviewers are unable to obtain interviews at about 3,200 of these units. This occurs when the occupants are not found at home after repeated calls or are unavailable for some other reason.

Sample Redesign. Since the introduction of the CPS, the Census Bureau has redesigned the CPS sample several times. These redesigns have improved the quality and accuracy of the data and have satisfied changing data needs. The most recent changes were phased in and implementation was completed in 1995.

Estimation Procedure. This survey's estimation procedure adjusts weighted sample results to agree with independent estimates of the civilian noninstitutional population of the United States by age, sex, race, Hispanic¹/non-Hispanic ancestry, and state of residence. The adjusted estimate is called the post-stratification ratio estimate. The independent estimates are calculated based on information from four primary sources:

- C The 1990 Decennial Census of Population and Housing.
- C An adjustment for undercoverage in the 1990 census.
- C Statistics on births, deaths, immigration, and emigration.
- C Statistics on the size of the armed forces.

The estimation procedure for the March supplement included a further adjustment so husband and wife of a household received the same weight. The independent population estimates include some, but not all, of undocumented immigrants.

ACCURACY OF THE ESTIMATES

A sample survey estimate has two types of error: sampling and nonsampling. The accuracy of an estimate depends on both types of error. The nature of the sampling error is known given the survey design. The full extent of the nonsampling error, however, is unknown.

Sampling Error. Since the CPS estimates come from a sample, they may differ from figures from a complete census using the same questionnaires, instructions, and enumerators. This possible variation in the estimates due to sampling error is known as Asampling variability.@

¹Hispanics may be of any race.

Nonsampling Error. All other sources of error in the survey estimates are collectively called nonsampling error. Sources of nonsampling error include the following:

- ⌄ Inability to get information about all sample cases (nonresponse).
- ⌄ Definitional difficulties.
- ⌄ Differences in interpretation of questions.
- ⌄ Respondent inability or unwillingness to provide correct information.
- ⌄ Respondent inability to recall information.
- ⌄ Errors made in data collection, such as recording and coding data.
- ⌄ Errors made in processing the data.
- ⌄ Errors made in estimating values for missing data.
- ⌄ Failure to represent all units with the sample (undercoverage).

Two types of nonsampling error that can be examined to a limited extent are nonresponse and coverage.

Nonresponse. The effect of nonresponse cannot be measured directly, but one indication of its potential effect is the nonresponse rate. For the April 2000 basic CPS, the nonresponse rate was 6.5%. The nonresponse rate for the supplement was an additional 7.4%, for a total supplement nonresponse rate of 13.4%.

Undercoverage. The concept of coverage in the survey sampling process is the extent to which the total population that could be selected for sample ~~Acovers~~ the survey's target population. CPS undercoverage results from missed housing units and missed people within sample households. Overall CPS undercoverage is estimated to be about 8 percent. CPS undercoverage varies with age, sex, and race. Generally, undercoverage is larger for males than for females and larger for Blacks and other races combined than for Whites.

The Current Population Survey weighting procedure uses ratio estimation whereby sample estimates are adjusted to independent estimates of the national population by age, race, sex and Hispanic ancestry. This weighting partially corrects for bias due to undercoverage, but biases may still be present when people who are missed by the survey differ from those interviewed in ways other than age, race, sex, and Hispanic ancestry. How this weighting procedure affects other variables in the survey is not precisely known. All of these considerations affect comparisons across different surveys or data sources.

A common measure of survey coverage is the coverage ratio, the estimated population before post-stratification divided by the independent population control. Table 2 shows CPS coverage ratios for age-sex-race groups for a typical month. The CPS coverage ratios can exhibit some variability from month to month. Other Census Bureau household surveys experience similar coverage.

Table 2. CPS Coverage Ratios

Age	Non-Black		Black		All People		
	M	F	M	F	M	F	Total
0-14	0.929	0.964	0.850	0.838	0.916	0.943	0.929
15	0.933	0.895	0.763	0.824	0.905	0.883	0.895
16-19	0.881	0.891	0.711	0.802	0.855	0.877	0.866
20-29	0.847	0.897	0.660	0.811	0.823	0.884	0.854
30-39	0.904	0.931	0.680	0.845	0.877	0.920	0.899
40-49	0.928	0.966	0.816	0.911	0.917	0.959	0.938
50-59	0.953	0.974	0.896	0.927	0.948	0.969	0.959
60-64	0.961	0.941	0.954	0.953	0.960	0.942	0.950
65-69	0.919	0.972	0.982	0.984	0.924	0.973	0.951
70+	0.993	1.004	0.996	0.979	0.993	1.002	0.998
15+	0.914	0.945	0.767	0.874	0.898	0.927	0.918
0+	0.918	0.949	0.793	0.864	0.902	0.931	0.921

Comparability of Data. Data obtained from the CPS and other sources are not entirely comparable. This results from differences in interviewer training and experience and in differing survey processes. This is an example of nonsampling variability not reflected in the standard errors. Therefore, caution should be used when comparing results from different sources.

A number of changes were made in data collection and estimation procedures beginning with the January 1994 CPS. The major change was the use of a new questionnaire. The questionnaire was redesigned to measure the official labor force concepts more precisely, to expand the amount of data available, to implement several definitional changes, and to adapt to a computer-assisted interviewing environment. The April child support supplement changes included refining the screening of potential respondents, restructuring the questionnaire to accommodate computerizing the survey, revising terminology that refers to types of child support agreements or awards, increasing the detail in questions about the amount of child support due, including overdue child support in the amount of child support due, and adding new questions on pass-through payments (child support collected for public assistance recipients by a state enforcement office, some of which passes through to recipients). Changes to the April CPS supplement in 1994 do not allow for comparisons with CPS data collected before that year.

Caution should also be used when comparing Hispanic estimates over time. No independent population control totals for people of Hispanic ancestry were used before 1985.

Based on the results of each decennial census, the Census Bureau gradually introduces a new sample design for the CPS. During this phase-in period, CPS data are collected from sample designs based on different censuses. While most CPS estimates were unaffected by this mixed sample, geographic estimates are subject to greater error and variability. Users should exercise caution when comparing estimates across years for metropolitan/nonmetropolitan categories.

A Nonsampling Error Warning. Since the full extent of the nonsampling error is unknown, one should be particularly careful when interpreting results based on small differences between estimates. Even a small amount of nonsampling error can cause a borderline difference to appear significant or not, thus distorting a seemingly valid hypothesis test. Caution should also be used when interpreting results based on a relatively small number of cases. Summary measures probably do not reveal useful information when computed on a base² smaller than 75,000.

For additional information on nonsampling error including the possible impact on CPS data when known, refer to

\$ Statistical Policy Working Paper 3, *An Error Profile: Employment as Measured by the Current population Survey*, Office of Federal Statistical Policy and Standards, U.S. Department of Commerce, 1978.

\$ Technical Paper 63RV, *Current Population Survey: Design and Methodology*, U.S. Census Bureau, U.S. Department of Commerce, 2002.

Standard Errors and Their Use. The sample estimate and its standard error enable one to construct a confidence interval. A confidence interval is a range that would include the average result of all possible samples with a known probability. For example, if all possible samples were surveyed under essentially the same general conditions and the same sample design, and if an estimate and its standard error were calculated from each sample, then approximately 90 percent of the intervals from 1.645 standard errors below the estimate to 1.645 standard errors above the estimate would include the average result of all possible samples.

A particular confidence interval may or may not contain the average estimate derived from all possible samples. However, one can say with specified confidence that the interval includes the average estimate calculated from all possible samples.

²subpopulation

Standard errors may be used to perform hypothesis testing. This is a procedure for distinguishing between population parameters using sample estimates. The most common type of hypothesis is that two population parameters are different. An example of this would be comparing the percentages of Whites with a college education to the percentage of Blacks with a college education.

Tests may be performed at various levels of significance. A significance level is the probability of concluding that the characteristics are different when, in fact, they are the same. For example, to conclude that two parameters are different at the 0.10 level of significance, the absolute value of the estimated difference between characteristics must be greater than or equal to 1.645 times the standard error of the difference.

The Census Bureau uses 90-percent confidence intervals and 0.10 levels of significance to determine statistical validity. Consult standard statistical textbooks for alternative criteria.

Estimating Standard Errors. To estimate the standard error of a CPS estimate, the Census Bureau uses replicated variance estimation methods. These methods primarily measure the magnitude of sampling error. However, they do measure some effects of nonsampling error as well. They do not measure systematic biases in the data due to nonsampling error. (Bias is the average of the differences, over all possible samples, between the sample estimates and the true value.)

Generalized Variance Parameters. Consider all the possible estimates of characteristics of the population that are of interest to data users. Now consider all the subpopulations such as racial groups, age ranges, etc. Finally, consider every possible comparison or ratio combination. The list would be completely unmanageable. Similarly, a list of standard errors to go with every estimate would be unmanageable.

Through experimentation, we have found that certain groups of estimates have similar relationships between their variances and expected values. We provide a generalized method for calculating standard errors for any of the characteristics of the population of interest. The generalized method uses generalized variance parameters for groups of estimates. These parameters are in Table 3 for April supplement data, and Table 4 for basic CPS monthly labor force estimates.

Standard Errors of Estimated Numbers. The approximate standard error, s_x , of an estimated num-

$$s_x = \sqrt{ax^2 + bx}$$

ber from this microdata file can be obtained using this formula:

Here x is the size of the estimate and a and b are the parameters in Table 2 or 3 associated with the particular type of characteristic. When calculating standard errors for numbers from cross-tabulations

involving different characteristics, use the factor or set of parameters for the characteristic which will give the largest standard error.

Illustration. Suppose you want to calculate the standard error and a 90 percent confidence interval of the number of unemployed females in the civilian labor force when the number of unemployed females in the civilian labor force is about 2,891,000. Use Formula (1) and the appropriate parameters from Table 3 to get:

Number, x	2,891,000
a parameter	-0.000018
b parameter	2,957
Standard error	92,000
90% conf. int.	2,740,000 to 3,042,000

$$s_x = \sqrt{-0.000018 \times 2,891,000^2 + 2,957 \times 2,891,000} = 92,000$$

where the standard error is calculated as

The 90-percent confidence interval is calculated as 2,891,000 " 1.645H92,000.

A conclusion that the average estimate derived from all possible samples lies within a range computed in this way would be correct for roughly 90 percent of all possible samples.

Standard Errors of Estimated Percentages. The reliability of an estimated percentage, computed using sample data from both numerator and denominator, depends on both the size of the percentage and its base. Estimated percentages are relatively more reliable than the corresponding estimates of the numerators of the percentages, particularly if the percentages are 50 percent or more. When the numerator and denominator of the percentage are in different categories, use the parameter from Table 2 or 3 indicated by the numerator.

The approximate standard error, $s_{x,p}$, of an estimated percentage can be obtained by using the following

$$s_{x,p} = \sqrt{\frac{b}{x} p (100 - p)}$$

formula:

Here x is the total number of people, families, households, or unrelated individuals in the base of the percentage, p is the percentage (0 # p # 100), and b is the parameter in Table 2 or 3 associated with the characteristic in the numerator of the percentage.

Illustration. In 1999, of the 11,499,000 custodial mothers in the United States, 32.2% were never married. Using the appropriate parameter from Table 2 and Formula (2) gives

Percentage, p	32.2
Base, x	11,499,000
b parameter	6,948
Standard error	1.15
90% conf. int.	30.3 to 34.1

where the standard error is calculated as

$$s_{x,p} = \sqrt{\frac{6,948}{11,499,000} \times 32.2 \times (100 - 32.2)} = 1.15$$

The 90-percent confidence interval is calculated as 32.2 " 1.645H1.15.

Standard Error of a Difference. The standard error of the difference between two sample estimates is approximately equal to

$$s_{xy} = \sqrt{s_x^2 + s_y^2}$$

where s_x and s_y are the standard errors of the estimates, x and y . The estimates can be numbers, percentages, ratios, etc. This will represent the actual standard error quite accurately for the difference between estimates of the same characteristic in two different areas, or for the difference between separate and uncorrelated characteristics in the same area. However, if there is a high positive (negative) correlation between the two characteristics, the formula will overestimate (underestimate) the true standard error.

For information on calculating standard errors for labor force data from the CPS which involve differences in consecutive quarterly or yearly averages, consecutive month-to-month differences in estimates, and consecutive year-to-year differences in monthly estimates see [Explanatory Notes and Estimates of Error: Household Data](#) in *Employment and Earnings*, a monthly report published by the Bureau of Labor Statistics.

Illustration. In 1999, of the 6,133,000 custodial mothers that were due child support, 2,818,000 or 45.9% received the full amount of child support due. Of the 658,000 custodial fathers that were due child support, 248,000 or 37.7% received the full amount of child support due.

	x	y	difference
Percentage, p	45.9	37.7	8.2

Base, x	6,133,000	658,000	-
b parameter	3,272	3,272	-
Standard error	1.15	3.42	3.61
90% conf. int.	44.0 to 47.8	32.1 to 43.3	2.3 to 14.1

$$s_{x-y} = \sqrt{1.15^2 + 3.42^2} = 3.61$$

The standard error of the difference is calculated as

The 90-percent confidence interval for the estimated difference between the households is calculated as $8.2 \pm 1.645(3.61)$. Because this interval does not include zero, we can conclude with 90-percent confidence that the percentage of custodial mothers due child support who received the full amount due is greater than the percentage of custodial fathers due child support who received the full amount due.

Accuracy of State Estimates. The redesign of the CPS following the 1980 census provided an opportunity to increase efficiency and accuracy of state data. All strata are now defined within state boundaries. The sample is allocated among the states to produce state and national estimates with the required accuracy while keeping total sample size to a minimum. Improved accuracy of state data was achieved with about the same sample size as in the 1970 design.

Since the CPS is designed to produce both state and national estimates, the proportion of the total population sampled and the sampling rates differ among the states. In general, the smaller the population of the state the larger the sampling proportion. For example, in Vermont approximately 1 in every 400 households is sampled each month. In New York the sample is about 1 in every 2,000 households. Nevertheless, the size of the sample in New York is four times larger than in Vermont because New York has a larger population.

Computation of Standard Errors for State Estimates. Standard errors for a state may be obtained by computing national standard errors, using formulas described earlier, and multiplying these by the appropriate f factor from Table 4. An alternative method for computing standard errors for a state is to multiply the a and b parameters in Table 2 or 3 by f^2 and then use these adjusted parameters in the standard error formulas.

Illustration. Suppose you want to calculate the standard error for the percentage of people 18 years old and over living in the state of New York who had completed a bachelor's degree or more. Suppose about 3,155,700 (23.1 percent) people had completed at least a bachelor's degree when there were about 13,684,000 people aged 18 and over living in New York. Following the first method mentioned above, use the appropriate parameter from Table 2 and Formula (2) to get:

Percentage, p	23.1
Base, x	13,684,000
b parameter	2,369
State factor, f	0.94
Standard error	0.55

Table 4 shows the f factor for New York to be 0.94. Thus, the standard error on the estimate of the percentage of people 18 and over in New York state who had completed college is approximately $0.55 \times 0.94 = 0.52$.

Following the alternative method mentioned above, obtain the needed state parameter by multiplying the parameter in table 3 by the f^2 factor in Table 4 for the state of interest. For example, for educational attainment for total or white in New York this gives

$b = 2,369 \times 0.89 = 2,108$. The standard error of the estimate of the percentage of people 18 and older in New York state who had completed college can then be found by using formula (2), the base of 13,684,000 and the new b parameter, 2,108. This gives a standard error of 0.52.

Technical Assistance. If you require assistance or additional information, please contact the Demographic Statistical Methods Division via e-mail at dsmd.source.and.accuracy@census.gov.

**Table 3. Parameters for Computation of Standard
Errors for Labor Force Characteristics: April 2000**

Characteristics	a	b
Civilian Labor Force, Employed, and Not in Labor Force		
<i>Total</i>	-0.000018	2,985
Men	-0.000033	2,764
Women	-0.000030	2,530
Both Sexes, 16-19 years	-0.000172	2,545
<i>White</i>	-0.000020	2,985
Men	-0.000037	2,767
Women	-0.000034	2,527
Both Sexes, 16-19 years	-0.000204	2,550
<i>Black</i>	-0.000125	3,139
Men	-0.000302	2,931
Women	-0.000183	2,637
Both Sexes, 16-19 years	-0.001295	2,949
Hispanic ancestry	-0.000206	3,896
Not in Labor Force (use only for Total, Total Men, and White)	0.000006	829
Agricultural Employment		
<i>Total or White</i>	0.000782	3,049
Men	0.000858	2,825
Women	-0.000025	2,582
Both Sexes, 16-19 years	-0.000025	2,582
<i>Black</i>	-0.000135	3,155
<i>Hispanic ancestry</i>		
Total or Women	0.011857	2,895
Men	0.015736	1,703
Both Sexes, 16-19 years	0.015736	1,703
Unemployment		
<i>Total or White</i>	-0.000018	2,957
<i>Black</i>	-0.000212	3,150
<i>Hispanic ancestry</i>	-0.000102	3,576

NOTE: These parameters are to be applied to basic CPS monthly labor force estimates.

For foreign-born and noncitizen characteristics for Total and White, the a and b parameters should be multiplied by 1.3. No adjustment is necessary for foreign-born and noncitizen characteristics for Blacks and Hispanics.

Multiply the a and b parameters by 1.5 when tabulating nonmetropolitan estimates.

Table 4. Standard Error Parameters for Child Support Characteristics

Characteristic	Total or White		Black		Hispanic	
	a	b	a	b	a	b
INCOME						
Persons	-0.000015	3,272	-0.000145	3,747	-0.000277	6,314
Families	-0.000014	2,988	-0.000127	3,263	-0.000241	5,498
POVERTY						
Persons Below the Poverty Level	-0.000051	13,840	-0.000391	13,840	-0.000711	23,324
NONINCOME						
Women/Men with Dependent Children Whose Fathers/Mothers are Absent						
Marital Status	-0.000025	6,948	-0.000282	9,981	-0.000513	16,821
SELECTED CHARACTERISTICS OF MEN AND WOMEN						
Education	-0.000015	3,159	-0.000139	3,573	-0.000179	4,069

Multiply the a and b parameters by 1.5 when tabulating nonmetropolitan estimates.

Table 5. State Factors

State	f	f ²	State	f	f ²
Alabama	1.00	1.01	Montana	0.45	0.20
Alaska	0.39	0.15	Nebraska	0.65	0.42
Arizona	0.98	0.97	Nevada	0.66	0.44
Arkansas	0.77	0.59	New Hampshire	0.62	0.38
California	1.14	1.29	New Jersey	0.91	0.82
Colorado	0.96	0.93	New Mexico	0.63	0.40
Connecticut	1.00	1.00	New York	0.94	0.89
Delaware	0.47	0.22	North Carolina	0.97	0.94
Dist. of Col.	0.40	0.16	North Dakota	0.40	0.16
Florida	0.98	0.97	Ohio	1.01	1.02
Georgia	1.18	1.40	Oklahoma	0.85	0.73
Hawaii	0.59	0.35	Oregon	0.93	0.86
Idaho	0.52	0.27	Pennsylvania	0.98	0.96
Illinois	1.00	1.00	Rhode Island	0.55	0.30
Indiana	1.17	1.38	South Carolina	1.00	1.01
Iowa	0.84	0.71	South Dakota	0.41	0.17
Kansas	0.81	0.65	Tennessee	1.16	1.34
Kentucky	0.96	0.92	Texas	1.10	1.21
Louisiana	0.97	0.95	Utah	0.66	0.43
Maine	0.61	0.37	Vermont	0.42	0.18
Maryland	1.17	1.38	Virginia	1.22	1.48
Massachusetts	0.90	0.81	Washington	1.21	1.47
Michigan	0.96	0.93	West Virginia	0.62	0.39
Minnesota	1.05	1.11	Wisconsin	1.11	1.23
Mississippi	0.80	0.64	Wyoming	0.35	0.12
Missouri	1.17	1.37			

Table 6. Region Factors

Characteristic	f	f ²
Northeast	0.92	0.85
Midwest	1.01	1.03
South	1.04	1.08
West	1.04	1.09